United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Research Division

RD Research Report
Number RD-96-01

September 1996

Within- and Between-Person Components of Variance for Nutrient Intakes in the United States, 1994

Patricia M. Guenther
Agricultural Research Service

Phillip S. Kott
National Agricultural Statistics Service

WITHIN- AND BETWEEN-PERSON COMPONENTS OF VARIANCE FOR NUTRIENT INTAKES IN THE UNITED STATES, 1994, by Patricia M. Guenther, Food Surveys Research Group, Beltsville Human Nutrition Research Center, Agricultural Research Service, and Phillip S. Kott, Research Division, National Agricultural Statistics Service, United States Department of Agriculture, Washington, D.C. 20250-2000, September 1996, Report Number RD-96-01.

ABSTRACT

The 1994-96 Continuing Survey of Food Intakes by Individuals was conducted to measure the kinds and amounts of foods consumed by the U.S. population. A national probability sample of individuals living in households was interviewed in their homes on two non-consecutive days and provided 24-hour recalls of their dietary intake. This report on components of variance provides design-based estimates of the within- and between-person coefficients of variation, based on data collected in 1994, for intakes of 26 nutrients and other dietary constituents for selected sex-age subpopulations.

KEY WORDS

Subpopulation; Design-based; Sample weight; Unbiased; Nearly unbiased.

This paper was prepared for limited distribution to the research community outside the U.S. Department of Agriculture.

ACKNOWLEDGMENT

The authors gratefully acknowledge Walter Willett, Harvard School of Public Health, whose request motivated this report.

TABLE OF CONTENTS

SUMMARY

The 1994-96 Continuing Survey of Food Intakes by Individuals was conducted to measure the kinds and amounts of foods consumed by the U.S. population. A national, stratified, multi-stage probability sample of individuals living in households was interviewed in their homes on two non-consecutive days and provided 24-hour recalls of their dietary intake. The survey was designed to provide an even spread of interviews across geographic regions, days of the week, and seasons of the year. The analytic response rate was 77 percent in 1994. Food intakes were linked to a food composition data base, and daily nutrient intakes for each respondent were calculated.

This report provides design-based estimates of the within-and between-person coefficients of variation, based on data collected in 1994, for intakes of 26 nutrients and other dietary constituents for selected sex-age subpopulations. (These descriptive statistics should not be confused with the common inferential statistic, the coefficient of variation of the estimated mean intake.) The within- and between-person coefficients of variation are of interest to nutrition scientists and others who study and assess dietary intake, and they can be particularly useful when planning clinical or epidemiologic studies. The estimation strategy is described in an appendix.

INTRODUCTION

The 1994-96 Continuing Survey of Food Intakes by Individuals (CSFII) was conducted by Westat, Inc., under contract to the U.S. Department of Agriculture (USDA) to measure the kinds and amounts of foods consumed by the U.S. population. A national, stratified, multi-stage probability sample of individuals living in households was interviewed in their homes on two non-consecutive days and provided 24-hour recalls of their dietary intake. The survey was designed to provide an even spread of interviews across seasons of the year and days of the week. The sample design called for the first dietary interview to be conducted on one of three specified days of the week, and the second was to be conducted 3 to 10 days later. The analytic response rate was 77 percent in 1994. Food intakes were linked to a food composition data base, and daily nutrient intakes for each respondent were calculated.

Results tabulated here were based on data collected in 1994 and weighted to adjust for nonresponse and unequal probabilities of selection and to equalize the number of interviews conducted in each of the four seasons and on each of the seven days of the week.

Additional details about the design and operation of the CSFII 1994-96 have been described elsewhere. The dataset and documentation are available to the public on CD-ROM (U.S. Department of Agriculture 1995).

This report develops estimates for the within- and between-person components of variance for intakes of nutrients and other dietary constituents for selected sex-age subpopulations. (These descriptive statistics should not be confused with the common inferential statistic, the variance of an estimated mean intake.) The within- and between coefficients of variation, presented in the following tables, are of interest to nutrition scientists and others who study and assess dietary intake, particularly because the within-person component is often very large compared to the between-person component (Beaton et al. 1979, Garn et al. 1978, Hegsted 1972, Marr 1971, Pao et al. 1985, Sempos et al. 1985, Subcommittee on Criteria for Dietary Evaluation 1986). This ratio is especially large for vitamin A, carotenes, and vitamin B_{12}. Coefficients of variation can be particularly useful in the planning of clinical or epidemiologic studies (Keys 1967, Willett 1990). The estimation strategy, described in the appendix, is designed-based rather than model-based. This approach results in estimates that are relatively free of model assumptions.

Estimates of within- and between-person coefficients of variation are presented by sex-age group (again, these statistics should not be confused with the coefficient of variation of an estimated mean intake). Estimates are not provided for individuals age 12 to 19 years because the sample size available from the 1994 CSFII was not large enough to provide reliable estimates of within- and between-person coefficients of variation for boys and girls separately, and the differences between food intakes by boys and girls were judged to be too large to support pooling of their data. Infants under 1 year of age are excluded because their intakes were judged to be too different from older children, and the number in the 1994 sample was too small to allow separate estimates. Fourteen breast-fed children were also excluded because their nutrient intakes were not determined.

Table 1. Estimated mean intakes of nutrients and other dietary components and within- and between-person coefficients of variation: *Children age 1 to 5 years*[1]
(sample size = 1,051; estimated population size = 20,609,536)

Nutrient	*Mean*	*Coefficient of variation (%)*	
		Within-person	*Between-person*
energy (kcal)	1442	27.5	25.8
protein (g)	52.3	33.2	28.1
total fat (g)	53.5	38.3	29.3
saturated fat (g)	20.5	40.9	30.2
monounsaturated fat (g)	20.3	40.9	32.1
polyunsaturated fat (g)	8.7	50.3	33.4
cholesterol (mg)	184	67.4	37.9
carbohydrate (g)	193	29.3	27.4
fiber (g)	9.7	42.7	34.5
vitamin A (μg RE)	778	101.8	50.2
carotenes (μg RE)	260	156.3	60.8
vitamin E (mg α-TE)	5.3	66.1	40.3
vitamin C (mg)	92	66.9	47.8
thiamin (mg)	1.21	36.9	27.1
riboflavin (mg)	1.71	35.4	31.2
niacin (mg)	14.6	37.4	32.7
vitamin B_6 (mg)	1.35	42.9	29.7
folate (μg)	196.	51.8	33.8
vitamin B_{12} (μg)	3.70	200.5	49.7
iron (mg)	11.5	46.9	32.1
magnesium (mg)	188.	30.8	28.1
calcium (mg)	789.	39.4	35.7
potassium (mg)	1951.	33.4	27.7
phosphorus (mg)	971.	31.1	28.3
zinc (mg)	7.5	43.9	30.8
copper (mg)	.8	51.7	29.5

[1]Excludes breast-fed children

SOURCE: USDA Continuing Survey of Food Intakes by Individuals, 1994

Table 2. Estimated mean intakes of nutrients and other dietary components and within- and between-person coefficients of variation: *Children age 6 to 11 years*
(sample size = 506; estimated population size = 23,157,637)

Nutrient	*Mean*	*Coefficient of variation (%)*	
		Within-person	*Between-person*
energy (kcal)	1850	28.9	22.3
protein (g)	65.6	33.6	26.
total fat (g)	68.1	39.1	23.1
saturated fat (g)	25.0	41.2	26.3
monounsaturated fat (g)	26.2	42.9	23.4
polyunsaturated fat (g)	11.9	53.5	16.8
cholesterol (mg)	213	62.1	43.6
carbohydrate (g)	251	30.7	25.4
fiber (g)	12.5	47.0	32.4
vitamin A (μg RE)	904	64.1	39.7
carotenes (μg RE)	312	146.3	72.6
vitamin E (mg α-TE)	6.5	55.8	30.1
vitamin C (mg)	98	68.9	46.4
thiamin (mg)	1.58	39.4	28.7
riboflavin (mg)	2.05	34.4	31.8
niacin (mg)	19.0	40.2	26.7
vitamin B_6 (mg)	1.65	43.5	35.2
folate (μg)	251	49.7	42.5
vitamin B_{12} (μg)	4.21	74.1	37.9
iron (mg)	13.8	46.9	38.2
magnesium (mg)	227	33.9	29.6
calcium (mg)	907	39.4	36.3
potassium (mg)	2251	34.0	29.3
phosphorus (mg)	1178	31.8	27.9
zinc (mg)	9.7	42.6	32.6
copper (mg)	1.0	38.4	23.3

SOURCE: USDA Continuing Survey of Food Intakes by Individuals, 1994

Table 3. Estimated mean intakes of nutrients and other dietary components and within- and between-person coefficients of variation: *Women age 20 to 59 years* (sample size = 1,093; estimated population size = 71,979,766)

Nutrient	*Mean*	*Coefficient of variation (%)*	
		Within-person	*Between-person*
energy (kcal)	1647	31.0	27.1
protein (g)	63.8	38.8	27.5
total fat (g)	61.4	44.3	32.9
saturated fat (g)	20.6	48.8	36.8
monounsaturated fat (g)	23.3	46.5	34.8
polyunsaturated fat (g)	12.8	62.8	32.1
cholesterol (mg)	220	78.0	37.1
carbohydrate (g)	209	33.5	29.3
fiber (g)	13.1	47.5	36.0
vitamin A (μg RE)	925	88.3	59.6
carotenes (μg RE)	465	170.3	71.3
vitamin E (mg α-TE)	7.2	76.8	54.7
vitamin C (mg)	86	84.2	58.9
thiamin (mg)	1.29	47.1	33.3
riboflavin (mg)	1.56	57.3	33.3
niacin (mg)	18.8	48.5	28.2
vitamin B_6 (mg)	1.48	53.7	32.3
folate (μg)	214	66.2	41.7
vitamin B_{12} (μg)	4.31	382.3	42.6
iron (mg)	12.4	54.2	34.4
magnesium (mg)	226	35.7	28.3
calcium (mg)	637	51.1	40.6
potassium (mg)	2259	35.2	27.8
phosphorus (mg)	1017	36.9	27.8
zinc (mg)	9.2	57.7	32.2
copper (mg)	1.0	72.8	26.1

SOURCE: USDA Continuing Survey of Food Intakes by Individuals, 1994

Table 4. Estimated mean intakes of nutrients and other dietary components and within- and between-person coefficients of variation: *Men age 20 to 59 years* (sample size = 1,082; estimated population size = 70,147,516)

Nutrient	*Mean*	*Coefficient of variation (%)*	
		Within-person	*Between-person*
energy (kcal)	2510	33.9	29.7
protein (g)	98.8	43.0	29.2
total fat (g)	95.7	45.8	30.8
saturated fat (g)	32.0	48.8	34.7
monounsaturated fat (g)	37.2	48.4	30.7
polyunsaturated fat (g)	19.1	63.8	30.9
cholesterol (mg)	342	67.7	39.2
carbohydrate (g)	298	36.0	33.3
fiber (g)	18.0	47.1	38.4
vitamin A (μg RE)	1070	99.4	57.1
carotenes (μg RE)	509	158.5	53.1
vitamin E (mg α-TE)	10.5	72.3	77.5
vitamin C (mg)	108	99.4	70.3
thiamin (mg)	1.94	46.4	36.1
riboflavin (mg)	2.24	45.6	37.3
niacin (mg)	29.1	46.9	34.0
vitamin B_6 (mg)	2.23	49.4	43.9
folate (μg)	300	60.0	53.5
vitamin B_{12} (μg)	5.92	122.4	37.2
iron (mg)	18.3	49.1	41.1
magnesium (mg)	327	38.0	31.3
calcium (mg)	888	53.3	41.8
potassium (mg)	3186	35.4	31.8
phosphorus (mg)	1494	37.4	28.9
zinc (mg)	14.3	62.3	37.9
copper (mg)	1.5	48.4	33.1

SOURCE: USDA Continuing Survey of Food Intakes by Individuals, 1994

Table 5. Estimated mean intakes of nutrients and other dietary components and within- and between-person coefficients of variation: *Women age 60 years and older* (sample size = 484; estimated population size = 23,221,414)

Nutrient	*Mean*	*Coefficient of variation (%)*	
		Within-person	*Between-person*
energy (kcal)	1401	32.0	25.9
protein (g)	57.8	38.6	23.6
total fat (g)	51.1	45.2	31.0
saturated fat (g)	16.9	48.3	36.4
monounsaturated fat (g)	19.1	48.8	30.2
polyunsaturated fat (g)	11.0	62.1	36.8
cholesterol (mg)	208	71.2	42.7
carbohydrate (g)	181	32.7	30.4
fiber (g)	13.9	39.8	39.7
vitamin A (μg RE)	1058	154.0	31.9
carotenes (μg RE)	524	151.2	48.6
vitamin E (mg α-TE)	6.3	65.0	33.3
vitamin C (mg)	92	59.2	60.4
thiamin (mg)	1.21	39.6	33.7
riboflavin (mg)	1.49	49.1	32.9
niacin (mg)	17.1	41.5	31.8
vitamin B_6 (mg)	1.44	44.6	36.0
folate (μg)	217	51.2	40.3
vitamin B_{12} (μg)	4.64	326.5	36.3
iron (mg)	11.56	49.2	36.9
magnesium (mg)	219	32.0	31.4
calcium (mg)	574	42.8	39.4
potassium (mg)	2255	29.2	28.6
phosphorus (mg)	919	33.6	28.9
zinc (mg)	8.0	50.6	26.4
copper (mg)	1.0	64.6	29.9

SOURCE: USDA Continuing Survey of Food Intakes by Individuals, 1994

Table 6. Estimated mean intakes of nutrients and other dietary components and within- and between-person coefficients of variation: *Men age 60 years and older* (sample size = 465; estimated population size = 17,516,913)

Nutrient	*Mean*	*Coefficient of variation (%)*	
		Within-person	*Between-person*
energy (kcal)	1979	28.3	28.4
protein (g)	79.4	34.8	24.1
total fat (g)	74.4	40.6	35.8
saturated fat (g)	24.8	44.7	41.8
monounsaturated fat (g)	28.7	42.4	38.5
polyunsaturated fat (g)	14.9	55.1	32.8
cholesterol (mg)	290	69.3	34.3
carbohydrate (g)	246	31.8	30.4
fiber (g)	18.4	42.3	42.5
vitamin A (μg RE)	1377	123.7	58.7
carotenes (μg RE)	642	117.1	94.9
vitamin E (mg α-TE)	9.2	58.7	77.9
vitamin C (mg)	105	62.9	60.0
thiamin (mg)	1.72	35.4	35.5
riboflavin (mg)	2.05	39.7	37.1
niacin (mg)	24.2	36.0	34.5
vitamin B_6 (mg)	2.04	37.6	42.5
folate (μg)	290	53.2	49.3
vitamin B_{12} (μg)	6.62	250.3	32.8
iron (mg)	17.3	41.9	50.6
magnesium (mg)	297	30.4	31.3
calcium (mg)	761	45.8	37.1
potassium (mg)	2944	30.6	27.9
phosphorus (mg)	1268	31.0	28.0
zinc (mg)	11.9	61.2	37.8
copper (mg)	1.3	55.4	31.0

SOURCE: USDA Continuing Survey of Food Intakes by Individuals, 1994

REFERENCES

U.S. Department of Agriculture, Agricultural Research Service (1995), *1994 Continuing Survey of Food Intakes by Individuals and 1994 Diet and Health Knowledge Survey*, CD-ROM, accession no. PB96-501010, National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161.

Beaton, G. H. Milner, J., Corey, P., McGuire, V., Cousins, M., Stewart, E., de Ramos, M., Hewitt, D., Grambsch, V., Kassim, N. & Little, J.A. (1979), Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. *American Journal of Clinical Nutrition* 32: 2546-2559.

Garn, S. M., Larkin, F. A., & Cole, P. (1978), The real problem with one-day records. *American Journal of Clinical Nutrition* 31: 1114-1116.

Hegsted, D. M. (1972), Problems in the use and interpretation of the Recommended Dietary Allowances. *Ecology of Food and Nutrition* 1: 255-265.

Keys, A. (1967), Dietary epidemiology. *American Journal of Clinical Nutrition* 20: 1151-1157.

Marr, J. W. (1971), Individual dietary surveys: purposes and methods. *World Review of Nutrition and Dietetics* 13: 105 - 164.

Pao, E. M., Mickle, S. J., & Burk, M. C. (1985), One-day and 3-day nutrient intakes by individuals. *Journal of the American Dietetic Association* 85: 313-324.

Pothoff, R. F., Woodbury, M.A., and Manton, K. G. (1992), "Equivalent sample size" and "equivalent degrees of freedom" refinements for inference using survey weights under superpopulation models. *Journal of the American Statistical Association* 87: 383-396.

Särndal, C. E. (1978), Design-based and model-based inference in survey sampling, *Scandinavian Journal of Statistics* 5: 27-52.

Sempos, C. T., Johnson, N. E., Smith, E. L., & Gilligan, C. (1985), Effects of Intraindividual and Interindividual Variation in Repeated Dietary Records. *American Journal of Epidemiology* 121: 120-130.

Subcommittee on Criteria for Dietary Evaluation (1986), *Nutrient Adequacy: Assessment Using Food Consumption Surveys*. National Academy Press, Washington, DC.

Willett, W. (1990), *Nutritional Epidemiology*, Oxford University Press, New York.

APPENDIX: Mathematical Formulae for Estimating the Within- and Between-Person Coefficients of Variation for Nutrient Intake Using the 1994 CSFII

The 1994-96 CSFII collected two days of food intake data from a stratified, multistage, random sample of individuals. Individuals from some demographic groups had a higher probability of being sampled than individuals from other groups; but from 1994-1996, each household resident in the United States had a positive probability of selection.

In this appendix, we focus on a particular nutrient and a particular subpopulation of interest (e.g., men age 60 and above), and show how the within- and between-person coefficients of variations for the daily intake of the nutrient within the subpopulation were calculated using the 1994 CSFII data.

Our goal was to compute estimates that were relatively simple to reproduce and relied on few assumptions. Accordingly, the approach to estimation taken was randomization- or design-based. In design-based inference, the targets of estimation (population means, within-person variances, etc.) are finite population parameters, which are analogous – but not identical – to the model parameters in a standard analysis of variance. Särndal (1978) provides a good introduction to the distinction between design- and model-based statistical inference. The sample design of the CSFII (see U.S. Department of Agriculture 1995) and the large CSFII sample sizes for the subpopulations of interest are, for the most part, conducive to model-free, design-based inference.

Let y_{it} denote the intake by individual i on day t. The mean daily intake for individual i in 1994 is defined as

$$\mu_i = \sum y_{it}/365,$$

where the summation is over the 365 days in 1994. The *within-person variance* for i is defined as

$$\sigma_{Wi}^2 = \sum (y_{it} - \mu_i)^2/364,$$

where the summation is again over all 365 days in 1994.

The 1994 CSFII provides two days of intake data for a sample of individuals. Suppose individual i is in the two-day CSFII sample, and let $y_{i(1)}$ and $y_{i(2)}$ denote his (her) two reported daily intakes. We will assume that these intakes are reported without error, and that each is an unbiased and independent estimator for μ_i (these are the only assumptions we make). As a result,

$$m_i = (y_{i(1)} + y_{i(2)})/2$$

is also an unbiased estimator for μ_i, while

$$s_{Wi}^2 = (y_{i(1)} - y_{i(2)})^2/2$$

is a nearly unbiased estimator for σ_{Wi}^2. This is because $E(y_{i(1)} - y_{i(2)}) = 0$, $Var(y_{i(1)} - y_{i(2)}) = Var(y_{i(1)}) + Var(y_{i(2)})$, and $\sigma_{Wi}^2 \approx Var(y_{i(1)}) = Var(y_{i(2)})$ under our assumptions. For future use, note that $E[(m_i - \mu_i)^2] = Var[(y_{i(1)} + y_{i(2)})/2] \approx \sigma_{Wi}^2/2$.

The *within-person variance for a population* P with N members is defined as

$$\sigma_W^2 = \sum_{i \in P} \sigma_{Wi}^2/N.$$

Suppose the 1994 CSFII contains a random, two-day sample of persons in P, which we denote S. As a practical matter, S is restricted to those individuals in the CSFII who provided two days of intake data.

A nearly unbiased estimator for σ_W^2 under our assumptions would be

$$s_W^2 = \sum_{i\in S} w_i s_{Wi}^2 / \sum_{i\in S} w_i,$$

where w_i is the two-day sampling weight associated with sample individual i.

The *mean intake* for population P is defined as

$$\mu = \sum_{i\in P} \mu_i / N,$$

and can be estimated in a nearly unbiased fashion under our assumptions by

$$m = \sum_{i\in S} w_i m_i / \sum_{i\in S} w_i.$$

The *between-person variance for the population* is defined as

$$\sigma_B^2 = \sum_{i\in P} (\mu_i - \mu)^2/(N - 1).$$

Observe that;

$$\begin{aligned}
E[(m_i - m)^2] &= E\{[(m_i - \mu_i) + (\mu_i - m)]^2\} \\
&= E[(m_i - \mu_i)^2] + \\
&\quad 2E[(\mu_i - m)(m_i - \mu_i)] + \\
&\quad E[(\mu_i - m)^2] \\
&= E[(m_i - \mu_i)^2] - \\
&\quad 2E[m(m_i - \mu_i)] + \\
&\quad E[(\mu_i - m)^2] \\
&= \sigma_{Wi}^2/2 - \sigma_{Wi}^2(w_i / \sum_{j\in S} w_j) \\
&\quad + E[(\mu_i - m)^2] \\
&\approx \sigma_{Wi}^2/2 + E[(\mu_i - m)^2] \\
&= \sigma_{Wi}^2/2 + (\mu_i - \mu)^2 + \\
&\quad 2(\mu_i - \mu)E(\mu - m) + \\
&\quad E[(\mu - m)^2],
\end{aligned}$$

assuming $w_i / \sum_{j\in S} w_j$ is small enough to be ignored.

In principle, a two-day sample of intakes could have been collected for every person in the population. Averaging across all individuals in P leads to;

$$\begin{aligned}
E[\sum_{i\in P} (m_i - m)^2/N] &= \sigma_W^2/2 + \\
&\quad \sum_{i\in P} (\mu_i - \mu)^2/N + \\
&\quad E[(\mu - m)^2] \\
&\approx \sigma_W^2/2 + \sigma_B^2,
\end{aligned}$$

where the final near equality holds when N is large and $E[(\mu - m)^2] << \sigma_W^2/2 + \sigma_B^2$. Both these conditions are satisfied for the subpopulations analyzed in the text.

Now $\sum_{i\in S} w_i(m_i - m)^2 / \sum_{i\in S} w_i$ is a nearly unbiased estimator for $\sum_{i\in P} (m_i - m)^2/N$. As a result, a nearly unbiased estimator for σ_B^2 is

$$s_B^2 = \sum_{i\in S} w_i(m_i - m)^2/\sum_{i\in S} w_i - s_W^2/2.$$

The estimated *within-person coefficient of variation* for the population is

$$cv_W = s_W/m * 100\%,$$

while the estimated *between-person coefficient of variation* for the population is

$$cv_B = s_B/m * 100\%.$$

Tables 1 through 6 report the sample sizes of the subpopulations of interest. In a conventional, model -based analysis of variance, the sample size provides an indirect measure of the accuracy of the computed estimates. Since the design-based estimates derived in this appendix are weighted, there is a consequent loss in accuracy even under ideal conditions. As a result, a more relevant indirect measure of accuracy than the sample

size is the *equivalent sample size,* which

Pothoff, Woodbury, and Manton (1992) define as:

$$n^* = (\sum_{i \in S} w_i)^2 / \sum_{i \in S} w_i^2.$$

The equivalent sample sizes (rounded to the nearest integer) for the six subpopulations of interest are listed below:

Children, 1 to 5 years	819
Children, 6 to 11 years	389
Men, 20 to 59 years	725
Women, 20 to 59 years	756
Men, 60 years and older	341
Women, 60 years and older	336